2A-02

Modeling HPC Job Mapping by Reconfiguring Free-Space Optics Links

Yao Hu, Ikki Fujiwara, Michihiro Koibuchi

National Institute of Informatics
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, JAPAN 101-8430
{huyao,ikki,koibuchi}@nii.ac.jp

Abstract

A number of small parallel applications run on a high-performance computing (HPC) system simultaneously. Job mapping becomes crucial to improve system utilization and user experience, because fragmentation of unused compute nodes could not be assigned for an incoming job with even a smaller size. Wireless supercomputers and datacenters with 60GHz radio or free-space optics (FSO) have been proposed so that a diverse application workload can be better supported by changing network topologies by swapping the endpoints of wireless links. In this study, we propose the use of FSO links for the purpose of improving job mapping. Furthermore, we evaluated a constrained use of partial FSO links. Our simulation results demonstrate that by directly connecting non-neighboring compute nodes, the FSO interconnection networks can achieve shorter average queuing length and time for all incoming jobs.

Keywords: Job mapping, interconnection networks, free-space optics, 60GHz wireless, high performance computing (HPC)

1. Introduction

Parallel jobs are composed of a set of processes or threads that cooperate to solve a computational problem. They are the mainstay of high-performance computing (HPC) [1]. Therefore, a crucial problem is how to allocate the resources of large-scale parallel platforms such as supercomputers and datacenters. The resource allocation problem is also known as the job mapping problem, on which we focus in this work to improve system utilization and user experience.

By a conventional mapping, each incoming job must be scheduled to run on available wire-connected neighboring processors. This is like packing embedded subgraphs. There is an obvious problem that we do not know future job arrivals when scheduling the current job. As a result, regardless of job scheduling policies used, conventional supercomputers with a fixed network topology usually have a low utilization and there are many cases where a new job cannot be mapped even though an enough number of compute nodes are not in use but they are disjoint.

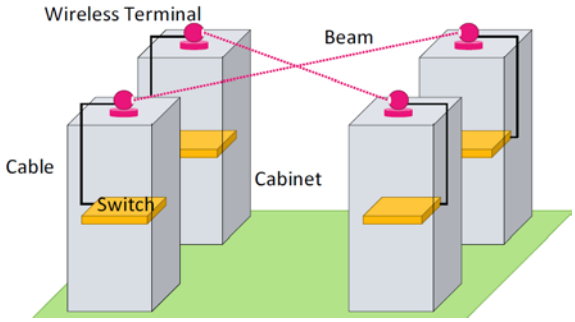

Figure 1. Wireless supercomputers and datacenters [2].

In this study, we attempt to solve the job mapping problem by using a wireless technology. Figure 1 shows an example of a wireless supercomputer and datacenter, which assumes that inter-cabinet links are wireless while intra-cabinet links are cables. An important property of this wireless system is that the link endpoints can be swapped on demand so as to increase the number of user jobs executed simultaneously [2]. Here, the term “wireless” indicates the technology of free-space optics (FSO) rather than the traditional 60GHz radio communication, since FSO has a similar bit error rate (BER) to the wired communication.

This paper presents the use of FSO links to replace wired links between compute nodes and investigates an efficient job mapping method on wireless supercomputers and datacenters. Evaluation results reveal that by swapping the endpoints of wireless links a large number of user jobs can be mapped simultaneously and the performance such as queuing length and time can thus be improved.

2. Methodology

We model job mapping problems with wired links and wireless links respectively, assuming that incoming jobs are scheduled to run on the system according to the time order.

As mentioned earlier, by a conventional job mapping on fixed network topologies, each job should be assigned to a set of neighboring unused compute

nodes that form a sub-topology with dilation 1. When inter-cabinet links are wireless, the corresponding two compute nodes can be directly connected. Thus such a sub-topology can easily be obtained with the help of wireless links. As shown in Fig. 2, obviously the three jobs cannot be assigned to the system simultaneously if the whole network topology maintains a 2-D mesh. However, swapping the endpoints of wireless links can successfully accommodate all the three jobs and thus improve the performance such as system utilization.

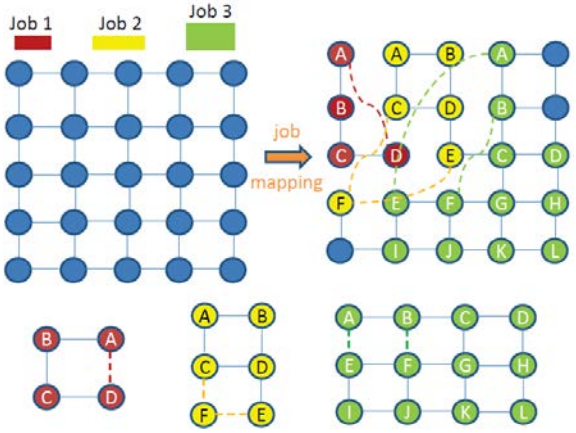

Figure 2. Job mapping over wireless links [3].

We also conservatively consider a partial provision of wireless links since they are expensive compared to wired cables. This would introduce difficulty in job mapping. Therefore, we designed a constrained job mapping algorithm [3] considering a part of the links are wireless while the remains are wired.

We use the Parallel Workloads Archive [1] as the data input in our evaluation, which collects real supercomputer traces of user jobs. By analyzing the trend of the job sizes, we found that the number of allocated processors varies widely for different jobs and most of the jobs request actually less than 100 processors. In each trace log file, we mainly use the data fields of Submit Time, Run Time and Number of Allocated Processors to simulate the job mapping in this work.

3. Evaluation

In this evaluation, we assume that the host network topology is 5-D torus. We use the early 196-hour traces from log file “UniLu Gaia” as the workload input. Since the degree of each node in 5-D torus is 5*2=10, the full FSO interconnection network enables 10 FSO links per node. In addition, we evaluate the case of the full wired interconnection network and partial FSO-introduced network where each node has the same portion of the FSO links over its entire links.

Figure 3 describes the average queuing length and time for all the jobs when we set different numbers of FSO links per node. The FSO links are used to replace the wired links so that any two non-neighboring compute nodes can be directly connected. Unsurprisingly, the full wired interconnection network (no FSO link) has the longest average queuing length and time, while the full FSO interconnection network (number of FSO links per node is 10) achieves the lowest value for both queuing length and time. In addition, the more use of partial FSO links also helps to improve the performance.

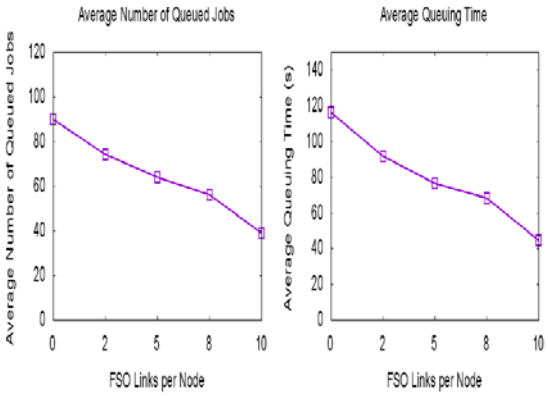

Figure 3. Average number of queued jobs (left) and queuing time (right) for job mapping over 5-D torus.

4. Conclusion

This paper proposes the use of FSO links in supercomputers and datacenters so that any two non-neighboring compute nodes can be directly connected. Evaluation results confirm that by swapping the endpoints of wireless links a large number of user jobs can be mapped simultaneously and the performance of full FSO interconnection networks such as average queuing length and time can thus be largely improved. In addition, the use of partial FSO links per node also helps to improve the performance by an efficient constrained job mapping algorithm.

Acknowledgement

This work was partially supported by SCOPE and JST CREST.

References

[1] D. G. Feitelson, “Workload Modeling for Computer Systems Performance Evaluation”, Cambridge University Press, March 2015.

[2] I. Fujiwara, M. Koibuchi, T. Ozaki, H. Matsutani, and H. Casanova, “Augmenting low-latency hpc network with free-space optical links”, the 21st International Conference on High-Performance Computer Architecture (HPCA), Feb 2015.

[3] Yao Hu, Ikki Fujiwara, and Michihiro Koibuchi, “Enabling Ideal Job Mapping on Wireless Supercomputers and Datacenters”, the 3rd International Workshop on Computer Systems and Architectures (CSA), Dec 2015.